Global Power Plant Database Project

Science one of the greatest invension is the electricity. It has many applications in our everyday life. It is used for lighting rooms, operating fans and household equipment such as electrical stoves, A/C and more. All of these give people warmth. Huge machines are operating in factories with the aid of electricity. Important items like food, fabric, paper and many other items are the result of electricity. The economic and overall infrastructure of a country depends on electricity.

Importing Libraries

In [ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Loading the DataSet

In [2]:
```
import pandas as pd
df=pd.read_csv(r"https://raw.githubusercontent.com/dsrscientist/dataset3/main/global_Power_plant_database.csv")
df.head()
```

Out[2]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	primary_fuel	other_fuel1	other_fuel2	...	geolocation_source	wepp_id	year_of_capacity_data	generation_gwh_201
0	IND	India	ACME Solar Tower	WRI1020239	2.5	28.1839	73.2407	Solar	NaN	NaN	...	National Renewable Energy Laboratory	NaN	NaN	Na
1	IND	India	ADITYA CEMENT WORKS	WRI1019881	98.0	24.7663	74.6090	Coal	NaN	NaN	...	WRI	NaN	NaN	Na
2	IND	India	AES Saurashtra Windfarms	WRI1026669	39.2	21.9038	69.3732	Wind	NaN	NaN	...	WRI	NaN	NaN	Na
3	IND	India	AGARTALA GT	IND0000001	135.0	23.8712	91.3602	Gas	NaN	NaN	...	WRI	NaN	2018.0	631.7779
4	IND	India	AKALTARA TPP	IND0000002	1800.0	21.9603	82.4091	Coal	Oil	NaN	...	WRI	NaN	2018.0	1668.2900

5 rows × 25 columns

In [34]:
```
df
```

Out[34]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	primary_fuel	other_fuel1	other_fuel2	...	geolocation_source	wepp_id	year_of_capacity_data	generation_gwh_
0	IND	India	ACME Solar Tower	WRI1020239	2.5	28.1839	73.2407	Solar	NaN	NaN	...	National Renewable Energy Laboratory	NaN	NaN	
1	IND	India	ADITYA CEMENT WORKS	WRI1019881	98.0	24.7663	74.6090	Coal	NaN	NaN	...	WRI	NaN	NaN	
2	IND	India	AES Saurashtra Windfarms	WRI1026669	39.2	21.9038	69.3732	Wind	NaN	NaN	...	WRI	NaN	NaN	
3	IND	India	AGARTALA GT	IND0000001	135.0	23.8712	91.3602	Gas	NaN	NaN	...	WRI	NaN	2018.0	631.7
4	IND	India	AKALTARA TPP	IND0000002	1800.0	21.9603	82.4091	Coal	Oil	NaN	...	WRI	NaN	2018.0	1668.2
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
903	IND	India	YERMARUS TPP	IND0000513	1600.0	16.2949	77.3568	Coal	Oil	NaN	...	WRI	NaN	2018.0	
904	IND	India	Yelesandra Solar Power Plant	WRI1026222	3.0	12.8932	78.1654	Solar	NaN	NaN	...	Industry About	NaN	NaN	
905	IND	India	Yelisirur wind power project	WRI1026776	25.5	15.2758	75.5811	Wind	NaN	NaN	...	WRI	NaN	NaN	
906	IND	India	ZAWAR MINES	WRI1019901	80.0	24.3500	73.7477	Coal	NaN	NaN	...	WRI	NaN	NaN	
907	IND	India	iEnergy Theni Wind Farm	WRI1026761	16.5	9.9344	77.4768	Wind	NaN	NaN	...	WRI	NaN	NaN	

908 rows × 25 columns

There are 908 rows and 25 columns. Each power plant contain information of it's plant capacity, gereration, fuel type, year of plant operation, electricity generation in gigawatt-hours in years from 2013 to 2017 and estimated generation in the following year etc.

In [35]:
```
df.columns
```

Out[35]:
```
Index(['country', 'country_long', 'name', 'gppd_idnr', 'capacity_mw',
       'latitude', 'longitude', 'primary_fuel', 'other_fuel1', 'other_fuel2',
       'other_fuel3', 'commissioning_year', 'owner', 'source', 'url',
       'geolocation_source', 'wepp_id', 'year_of_capacity_data',
       'generation_gwh_2013', 'generation_gwh_2014', 'generation_gwh_2015',
       'generation_gwh_2016', 'generation_gwh_2017', 'generation_data_source',
       'estimated_generation_gwh'],
      dtype='object')
```

In [3]:
```
df.shape
```

Out[3]:
```
(908, 25)
```

All the column labels from datafram can obtain by colums method.

Data Preparation & Cleaning

In [36]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 908 entries, 0 to 907
Data columns (total 25 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   country                   908 non-null    object 
 1   country_long              908 non-null    object 
 2   name                      908 non-null    object 
 3   gppd_idnr                 908 non-null    object 
 4   capacity_mw               908 non-null    float64
 5   latitude                  862 non-null    float64
 6   longitude                 862 non-null    float64
 7   primary_fuel              908 non-null    object 
 8   other_fuel1               199 non-null    object 
 9   other_fuel2               1 non-null      object 
 10  other_fuel3               0 non-null      float64
 11  commissioning_year        528 non-null    float64
 12  owner                     342 non-null    object 
 13  source                    908 non-null    object 
 14  url                       908 non-null    object 
 15  geolocation_source        889 non-null    object 
 16  wepp_id                   0 non-null      float64
 17  year_of_capacity_data     520 non-null    float64
 18  generation_gwh_2013       384 non-null    float64
 19  generation_gwh_2014       401 non-null    float64
 20  generation_gwh_2015       425 non-null    float64
 21  generation_gwh_2016       437 non-null    float64
 22  generation_gwh_2017       443 non-null    float64
 23  generation_data_source    450 non-null    object 
 24  estimated_generation_gwh  0 non-null      float64
dtypes: float64(13), object(12)
memory usage: 177.5+ KB
```

In [8]:
```
df.head()
```

Out[8]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	primary_fuel	other_fuel1	other_fuel2	...	geolocation_source	wepp_id	year_of_capacity_data	generation_gwh_201
0	IND	India	ACME Solar Tower	WRI1020239	2.5	28.1839	73.2407	Solar	NaN	NaN	...	National Renewable Energy Laboratory	NaN	NaN	Na
1	IND	India	ADITYA CEMENT WORKS	WRI1019881	98.0	24.7663	74.6090	Coal	NaN	NaN	...	WRI	NaN	NaN	Na
2	IND	India	AES Saurashtra Windfarms	WRI1026669	39.2	21.9038	69.3732	Wind	NaN	NaN	...	WRI	NaN	NaN	Na
3	IND	India	AGARTALA GT	IND0000001	135.0	23.8712	91.3602	Gas	NaN	NaN	...	WRI	NaN	2018.0	631.7779
4	IND	India	AKALTARA TPP	IND0000002	1800.0	21.9603	82.4091	Coal	Oil	NaN	...	WRI	NaN	2018.0	1668.2900

5 rows × 25 columns

In [9]:
```
df.tail()
```

Out[9]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	primary_fuel	other_fuel1	other_fuel2	...	geolocation_source	wepp_id	year_of_capacity_data	generation_gwh_
903	IND	India	YERMARUS TPP	IND0000513	1600.0	16.2949	77.3568	Coal	Oil	NaN	...	WRI	NaN	2018.0	
904	IND	India	Yelesandra Solar Power Plant	WRI1026222	3.0	12.8932	78.1654	Solar	NaN	NaN	...	Industry About	NaN	NaN	
905	IND	India	Yelisirur wind power project	WRI1026776	25.5	15.2758	75.5811	Wind	NaN	NaN	...	WRI	NaN	NaN	
906	IND	India	ZAWAR MINES	WRI1019901	80.0	24.3500	73.7477	Coal	NaN	NaN	...	WRI	NaN	NaN	
907	IND	India	iEnergy Theni Wind Farm	WRI1026761	16.5	9.9344	77.4768	Wind	NaN	NaN	...	WRI	NaN	NaN	

5 rows × 25 columns

In [5]:
```
df.name
```

Out[5]:
```
0                   ACME Solar Tower
1               ADITYA CEMENT WORKS
2          AES Saurashtra Windfarms
3                        AGARTALA GT
4                       AKALTARA TPP
                    ...
903                     YERMARUS TPP
904     Yelesandra Solar Power Plant
905     Yelisirur wind power project
906                      ZAWAR MINES
907          iEnergy Theni Wind Farm
Name: name, Length: 908, dtype: object
```

In [11]:
```
df.describe()
```

Out[11]:

	capacity_mw	latitude	longitude	other_fuel3	commissioning_year	wepp_id	year_of_capacity_data	generation_gwh_2013	generation_gwh_2014	generation_gwh_2015	generation_gwh_2016
count	908.000000	862.000000	862.000000	0.0	528.000000	0.0	520.0	384.000000	401.000000	425.000000	437.000000
mean	321.046378	21.196189	77.447848	NaN	1996.876894	NaN	2018.0	2304.059202	2420.393316	2414.072373	2453.936292
std	580.221767	6.248627	4.907260	NaN	17.047817	NaN	0.0	3794.767492	4013.558173	4183.203199	4152.038216
min	0.000000	8.168900	68.644700	NaN	1927.000000	NaN	2018.0	0.000000	0.000000	0.000000	0.000000
25%	16.837500	16.771575	74.258975	NaN	1988.000000	NaN	2018.0	244.458088	223.650436	174.174750	187.193669
50%	60.000000	21.778300	76.719250	NaN	2000.000000	NaN	2018.0	797.063475	805.760000	701.027250	716.728350
75%	388.125000	25.516375	79.441475	NaN	2011.250000	NaN	2018.0	2795.021500	3034.575000	3080.000000	3263.483000
max	4760.000000	34.649000	95.408000	NaN	2018.000000	NaN	2018.0	27586.200000	28127.000000	30539.000000	30015.000000

In [12]:
```
df.isnull()
```

Out[12]:

	country	country_long	name	gppd_idnr	capacity_mw	latitude	longitude	primary_fuel	other_fuel1	other_fuel2	...	geolocation_source	wepp_id	year_of_capacity_data	generation_gwh_2013	g
0	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
1	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
2	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
3	False	False	False	False	False	False	False	False	True	True	...	False	True	False	False	
4	False	False	False	False	False	False	False	False	False	True	...	False	True	False	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
903	False	False	False	False	False	False	False	False	False	True	...	False	True	False	True	
904	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
905	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
906	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	
907	False	False	False	False	False	False	False	False	True	True	...	False	True	True	True	

908 rows × 25 columns

In [14]:
```
df.isnull().sum()
```

Out[14]:
```
country                       0
country_long                  0
name                          0
gppd_idnr                     0
capacity_mw                   0
latitude                     46
longitude                    46
primary_fuel                  0
other_fuel1                 709
other_fuel2                 907
other_fuel3                 908
commissioning_year          380
owner                       566
source                        0
url                           0
geolocation_source           19
wepp_id                     908
year_of_capacity_data       388
generation_gwh_2013         524
generation_gwh_2014         507
generation_gwh_2015         483
generation_gwh_2016         471
generation_gwh_2017         465
generation_data_source      458
estimated_generation_gwh    908
dtype: int64
```

In [15]:
```
df.isnull().sum().sum()
```

Out[15]:
```
8693
```

Exploratory Analysis and Visualization

Exploratory Data Analysis (EDA) is a process of describing the data by means of statistical and visualization techniques in order to bring important aspects of that data into focus for further analysis. This involves inspecting the dataset from many angles, describing & summarizing it without making any assumptions about its contents.

In [ ]:
```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (12, 8)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

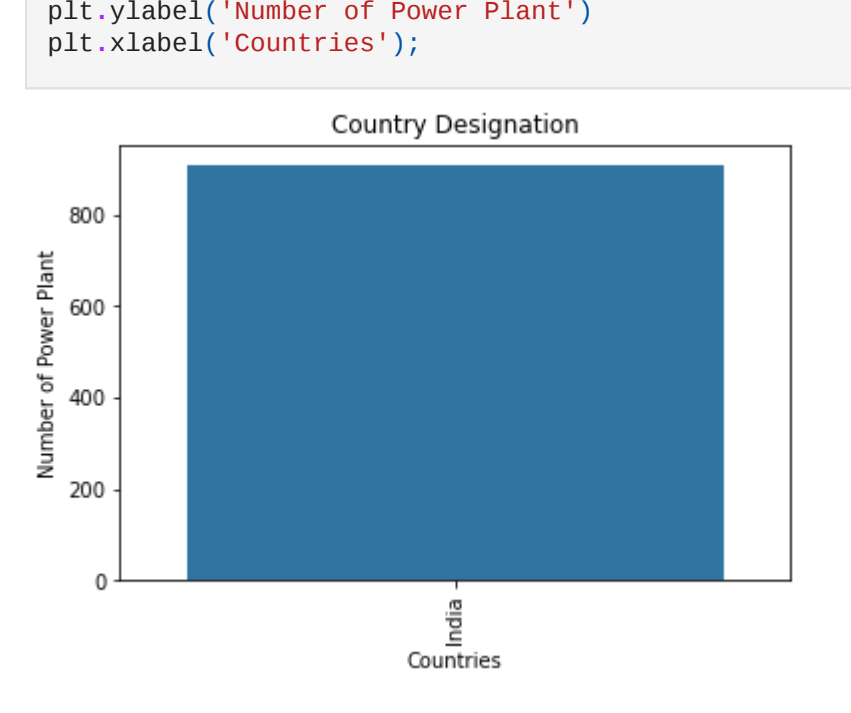
In [21]:
```
df.country_long.nunique()
```

Out[21]:
```
1
```

In [22]:
```
countries_plant = df.country_long.value_counts().head(20)
countries_plant
```

Out[22]:
```
India    908
Name: country_long, dtype: int64
```

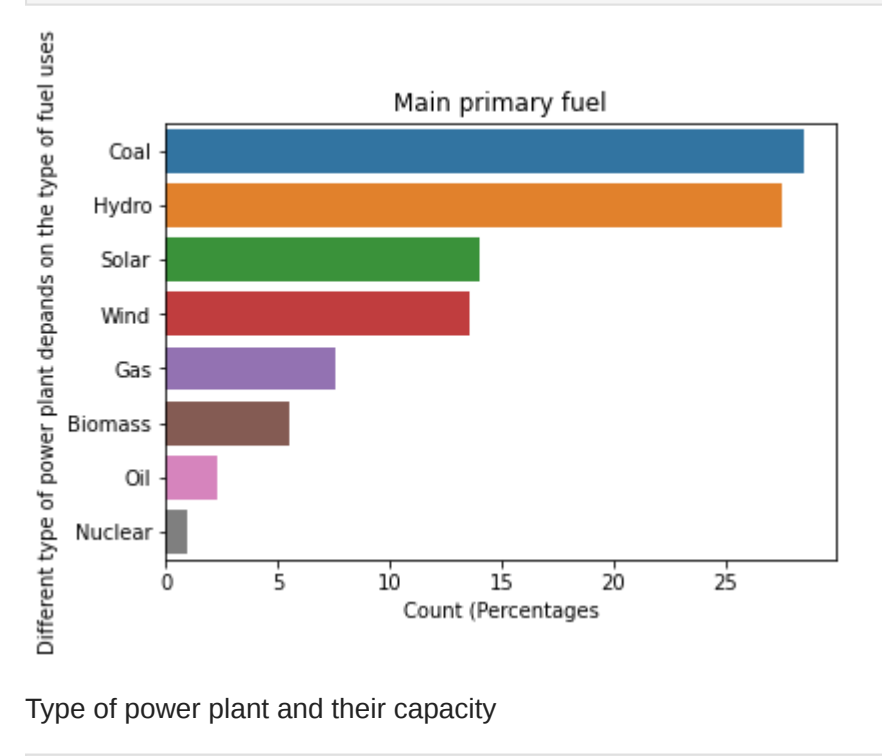
In [24]:
```
import matplotlib.pyplot as plt
sns.barplot(x = countries_plant.index, y = countries_plant)
plt.xticks(rotation = 90)
plt.title('Country Designation')
plt.ylabel('Number of Power Plant')
plt.xlabel('Countries');
```



In [25]:
```
main_primary_fuel = df.primary_fuel.value_counts() * 100 / df.primary_fuel.count()
main_primary_fuel
```

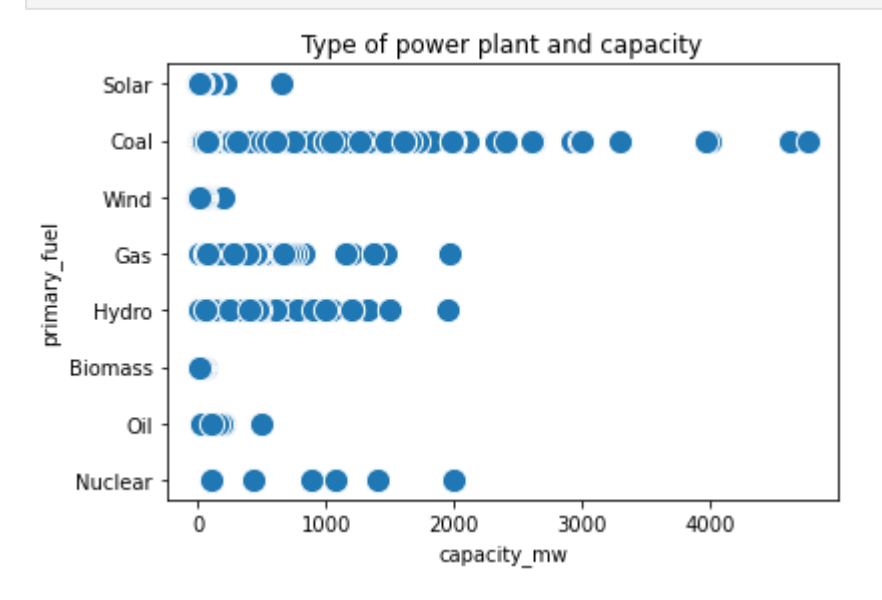
Out[25]:
```
Coal       28.524229
Hydro      27.533040
Solar      13.986784
Wind       13.546256
Gas         7.599119
Biomass     5.506608
Oil         2.312775
Nuclear     0.991189
Name: primary_fuel, dtype: float64
```

In [28]:
```
sns.barplot(x = main_primary_fuel, y = main_primary_fuel.index)
plt.title('Main primary fuel')
plt.xlabel('Count (Percentages');
plt.ylabel('Different type of power plant depands on the type of fuel uses');
```



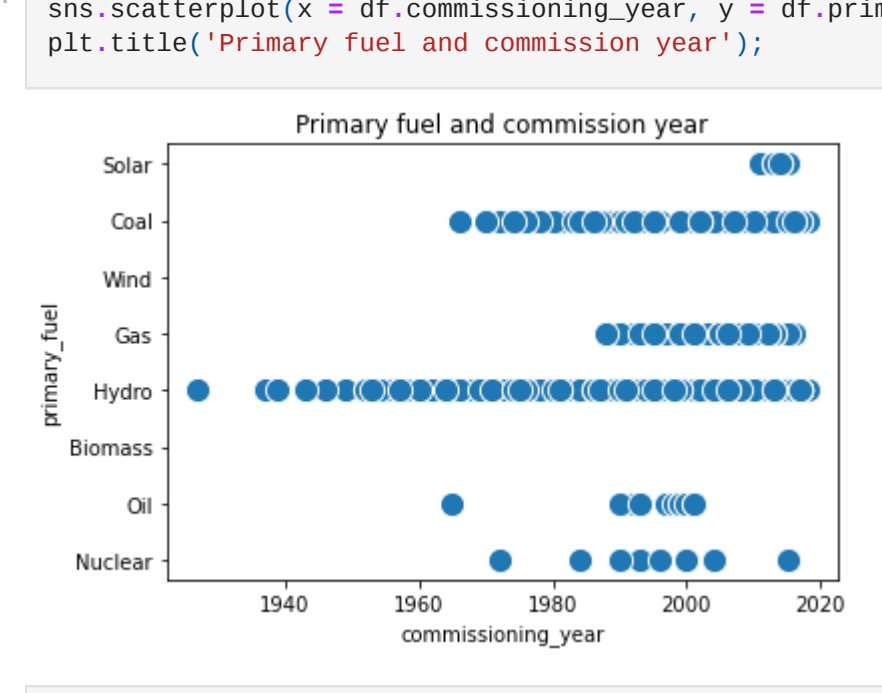
Type of power plant and their capacity

In [29]:
```
sns.scatterplot(x = df.capacity_mw, y = df.primary_fuel, s = 150)
plt.title('Type of power plant and capacity');
```



Different type of primary fuel based power plant and their year of going to the first operation.

In [32]:
```
sns.scatterplot(x = df.commissioning_year, y = df.primary_fuel, s = 150);
plt.title('Primary fuel and commission year');
```



In [33]:
```
df.primary_fuel.unique()
```

Out[33]:
```
array(['Solar', 'Coal', 'Wind', 'Gas', 'Hydro', 'Biomass', 'Oil',
       'Nuclear'], dtype=object)
```

Conclusion

This analysis gives us an insight into the depth of power plants around the world. How the world produces one of the most important elements, as the country's economic and overall infrastructure depends on electricity.

The world is facing global warming, and pollution from power plants is one of the reasons for this. We need to control pollution and urge on countries around the world to build more renewable or green energy power plants.

THANK YOU